DOCUMENT RESUME

ED 064 574 AC 012 647

AUTHOR Bhola, H. S.
TITLE Making Evaluation "Operational" in Functional Literacy Programs.
NOTE 36p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Data Collection; *Decision Making; *Evaluation Methods; *Functional Illiteracy; Input Output Analysis; *Program Evaluation; *Systems Approach
IDENTIFIERS UNESCO; *World Functional Literacy Program

ABSTRACT

Steps to be taken to "operationalize" evaluation, i.e., to make evaluation work, are discussed and applied to the Unesco program of functional literacy. Evaluation is seen as an important social concern, as well as a field of study and an area of competence. Four types of evaluation applied to the Unesco program are context evaluation, input evaluation, process evaluation, and output evaluation. The steps taken in operationalization are: (1) developing a generalized system model for the program; (2) designing an appropriate evaluation system; (3) designing a data system. It is stressed that evaluation is a group task from planning through completion, with the evaluator fulfilling his special technical role. A data system for functional literacy is shown to include both numerical information and documentation, with most of the data relating to four social configurations--individuals, groups, institutions, and sub-cultures--but with some of the data relating to tools, techniques, and artifacts that the four social configurations worked with to create development. In conclusion, it is pointed out that there is no complete evaluation system, rather, each project must build its evaluation system in terms of its needs, priorities, and resources. (DB)

ED 064574

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

FILMED FROM BEST AVAILABLE COPY

MAKING EVALUATION OPERATIONAL IN FUNCTIONAL LITERACY PROGRAMS

H.S. Bhola[1]

Evaluation is comparatively new both as a social concern and as an area of technical competence. In recent years, happily, there has been considerable progress. There is much greater conceptual clarification: we are beginning to understand the nature of evaluation; and how, in its approach and in its objectives, evaluation is different from both research and supervision. However, while within the formal school systems evaluation has already made the first breakthrough, problems remain in the area of evaluation of broad impact programs like functional literacy where program effects are many, some effects are delayed, and some rather diffused.

Functional literacy is one approach to social change. It seeks not only to teach adults to read and write and to make them more productive, but also it seeks, through an integrated program of work-oriented literacy, to touch their total lives, to transform their social, economic and value structures. Therefore, to be able to evaluate functional literacy programs (or any literacy program that has the objectives of broader social, economic or cultural impact) is not so easy. There is not much available to go by. Understandably, many evaluation programs in literacy, extension, and development never get off the ground; others remain sporadic, without integrity or direction, and, therefore, of no use to decision-makers within projects who must be served by such evaluation programs.

---

[1]Dr. H.S. Bhola previously a Unesco Expert on the Work-Oriented Adult Literacy Pilot Project in Tanzania is now an Associate Professor of Education, School of Education, Indiana University, Bloomington, Indiana, USA, 47401. While at the University he has continued a consultancy relationship with Unesco's Experimental World Functional Literacy Program advising in the area of both evaluation and training.

AC012647

In the following we outline the steps to be taken to operationalize evaluation: to make evaluation work. We suggest, that is, how the decision to evaluate a program must be worked through to where the decision can be implemented. What does one do first? What does one do next, and next after that, to design and to establish a fully functioning and useful evaluation system?

We begin by suggesting that it is impossible to design an evaluation program without first understanding the program of work: its objectives, its parts, its processes. An Evaluation System is a shadow cast by the Program System. One can never design an evaluation system without first understanding the design of the program system.

Programs can have broad impact on all the various aspects of the lives of people, both clients and change agents, and on communities and institutions involved. Again, such impact can be immediate or delayed, focussed or diffused. Different approaches and technqiues could be used to achieve those objectives. Thus there is, theoretically speaking, an infinite number of things to count and characteristics to measure and thus a never-ending list of evaluation questions that could be asked. An evaluation program could be designed that would take decades to implement. It is important, therefore, to be appropriately selective; which means that it is important to have an evaluation policy to be able to derive an evaluation system from a program system based on priorities and significant needs of decision-makers. Once an evaluation system has been established, a supportive data system can be created to serve the evaluation system which in turn will serve the program system.

Since the Unesco/UNDP Experimental World Functional Literacy Program is currently the most significant literacy program with deep interest in

evaluation we include a brief discussion of some special problems that have emerged in the course of creating a mata evaluation system--an evaluation system covering many project evaluation systems comprising the total Unesco program. This discussion has some useful lessons for those interested in creating regional or national evaluation and data systems in education and development.

I

Evaluation as a Social Concern

Evaluation, we have suggested, is comparatively new both as a social concern and as an area of technical competence. Programs and projects of development are today being questioned with regard to their stated goals, their hidden objectives, for choice of strategies that may be value-laden, for performance and results--both intended and unintended. The erstwhile helpless recipients of programs attempting to assert their newly acquired political rights, or simply, to celebrate their new social awareness want to find out about who is getting what, at what cost, and at whose cost. How well are they doing--those planners, designers, developers, and change agents working on change programs? Are they effective? Are they making the right decisions? Couldn't they do better than they are doing now? There is a political demand that the educational process, the extension process, the development process show some tangible results and that they deliver what is supposed to be delivered, where it is supposed to be delivered. The concern with effectiveness and the demand for accountability can be met only through evaluation.

In the developing world, as in the developed, there is a burgeoning

interest in evaluation. One reason for this increased interest in evaluation may be rather spurious, but another seems more genuine. Donor governments and agencies are often selling evaluation as part of the technical assistance package--even when there may be available neither the understandings nor the resources for undertaking such evaluations. In other cases institutions in the Third World may be getting on the evaluation bandwagon to be with it all.

But, as we suggested earlier, there is a genuine reason why evaluation is becoming more and more the usual thing to do both in the developed and the underdeveloped world. It is the <u>competition</u> for scarce resources. As we learn more and more about the use of social and behavioral sciences in inventing new futures for individuals, groups, and communities, we are using many different strategies and many different program approaches. Funding agencies and planning commissions when confronted with advocates of different approaches to effective fulfillment of the same objectives want to see comparative data on outputs received for the inputs made. This then is what has made evaluation an important social concern: politicization of recipient communities and societies; and the competition for scarce resources.

<u>Evaluation as an Area of Competence</u>

Evaluation is new also as a field of study and an area of competence. We had confused it, on the one hand, with research and, on the other, with supervision. Things are becoming clearer now.

We thought evaluation was research since evaluators collected scores which the researchers also did; evaluators made tables and graphs like the researchers also made; evaluators used tests which researchers also used. A carpenter may use the same tools for making a bridal bed or a coffin, but that does not

mean that the bridal bed and the coffin are the same thing. The same professional worker may be able to play two roles--the researcher role and the evaluator role. In both these roles--at one time as researcher, at another time as evaluator--he may use the same tools, like tests and questionnaires and interviews, but that does not make evaluation and research the same thing. They may infact, in functional terms, be as different as the bridal bed and the coffin. You lie down in both, but with differing objectives and surrounded by different setsof circumstances.

We now seem to know, for instance, that while both the researcher and evaluator may come from academic background of the university, the researcher will have scholarly interests while the evaluator will be more action-oriented. The researcher will most often choose the research question that might interest him personally, but the evaluator will pick up his evaluation question in consultation with the program team for whom he would be doing the evaluation. The evaluator would be responsive to the group needs, to the needs of the project. Again, while the researcher will pick up the area which may be pragnant with theoretical significance and might bring about a breakthrough in that whole field of intellectual endeavor; the evaluator will pick up a subject of practical significance to the project; it will be his desire to invent a solution to the problem with which his project might be faced with at that time. Thus, while the researcher and evaluator would use the same tools, the researcher would be more interested in perfect design, treatments without contamination, and statistical significance; the evaluator would be pragmatic, using the best design possible under the circumstances, inviting contamination because that is how things are out there in a nonlaboratory situation, and would be interested not in statistical significance, but with practical significance.

Again, the researcher may have more time on his hands to follow his interests while the evaluator may not. The evaluator must produce some feedback by such and such time, for the team to be able to make a new decision or to put more confidence in a decision already made. He cannot ask for more time since there may be no more time.

Lastly, the researcher and the evaluator are themselves evaluated differently in terms of their performance. The researcher would be evaluated in terms of whether he made an original contribution to human knowledge and produced publishable research. The evaluator would be evaluated in terms of whether he was able to provide the information needed by the decision makers in a specific project, to make the specific program decisions, that had to be made; and whether he was able to produce the information needed to measure outputs for the inputs applied.

It must be said in the end that a good evaluation study may sometimes serve both immediate project needs and at the same time make a contribution to human knowledge thereby becoming what we have been describing as research. This may happen, but the distinction that we have made between research and evaluation is valid and should be understood.

Evaluation, again, is <u>not</u> supervision; certainly not in the traditional meaning of supervision as policing subordinate officials in the field or within the organization. Evaluation, as we will see, produces feedback information on what is happening to the clients of the programs launched in the field, information on how the clients are perceiving the program, what attitudes and skills they are learning. It is usual in programs of innovation diffusion, planned change, and development, to find a big gap between intentions and actualities. <u>Between the intended and the actual there is always a parallax</u>.

This can be stated as a law. However, program administrators may often fail to analyze the reasons for discrepancies between their intentions and actualities in the field in terms of inappropriate models they used, of communication barriers, or cultural experiences of individuals, groups, and communities. They may consider it the "fault" of the field staff that did not work harder to achieve the objectives.

That is a very naive view of social and cultural dynamics. While breakdowns may occur because a field worker was not where he was supposed to be and did not do what he was supposed to do, causes of discrepancies between program intentions and actualities in the field seldom lie in staff irresponsibility. Evaluation may indeed discover some staff irresponsibility, but that is not its objective. Its objective is to make analyses, not allegations; to find causes, not culprits.

This distinction between evaluation and supervision must be understood not only by project administrators and planners, but should be widely shared among all workers and staff. If this is not done, the subordinate staff would feel afraid, and to cover their "faults" would cover reality, would falsify data, making evaluation a mere ritual, pushing planners in a world of fantasy and, ultimately, of hard knocks.

II

Evaluation: Definition and Types

In the preceding, we have tried to state what evaluation is not. Let us now say what evaluation is. How evaluation might be defined.

We include in the following a definition of evaluation suggested by Guba and Stufflebeam[1] who have made important contributions to the theory and method-

[1]Egon G. Guba and Daniel L. Stufflebeam, Evaluation: The Process of Stimulating, Aiding, and Abetting Insightful Action. (Monograph Series in Reading Education, Indiana University, Bloomington), 1970.

ology of evaluation:

> "Evaluation is the process of obtaining and providing useful information for making educational decisions."

Decisions is a key word in the definition. Evaluation assists in the making of decisions; that is its total justification. This assistance, it provides, by obtaining and providing to the decision makers information based on data.

The System Metaphor

We have already used the word system in talking about the program system, the evaluation system, and the data system. What is a system? A system, simply, is a whole with interdependent parts. The system metaphor points out that in life everything is connected with everything else. If one thing happens to one part, other parts are always in some ways effected. Human body is a system. If one thing happens to one part of the body, the whole body is indeed effected. Again, the human body as a system may be seen to be a System made up of overlapping systems: optical system, nervous system, breathing system, digestive system, etc.

The system metaphor is useful because all systems can be described in terms of four system parameters: context, input, process, and output. Every living and social system within the human cultural experience could be described in terms of these four parameters. Guba and Stufflebeam use the system metaphor to describe types of evaluation. Since any system can be described in terms of four parameters--context, input, process, and output--there can be four types of evaluation: context evaluation, input evaluation, process evaluation, and output evaluation.[1]

---

[1] Guba and Stufflebeam, op. cit., page 26-28.

By defining these four parameters: context, input, process, and output, we can define the four types of evaluation: context evaluation, input evaluation, process evaluation, and output evaluation.

Context, quite simply, is what surrounds the system, its environment, the flux from which the system in question has been separated by a system boundary. Context evaluation is thus evaluation related to analyzing, defining, and measuring the environment of the program system. Context evaluation helps in planning decisions. It talks of what needs might exist. What kind of climate for change may be experienced?

Input is what you put in a program system--men, materials, tools. Input evaluation thus is concerned with determining alternatives in the utilization of resources to achieve program goals.

Process deals with the question 'How'?, with procedures, arrangements, formations, extension, education. Process evaluation is thus concerned with designing procedural strategies and comparing the effectiveness of different approaches to instruction, extension, animation, and organization.

Lastly, output is what is put out--the product of the system. Therefore, output evaluation is concerned with determining the quantity and the quality of the product of the program, and whether there were any unintended social or economic by-products of the program.

Sometimes we have made the mistake of thinking that output evaluation is the only type of evaluation. That is not so. Output evaluation is only one out of the four types of evaluation and the four types complement each other. Context evaluation helps in planning, and programs do need to be properly planned if they have to succeed. Input evaluation helps in the deployment of men and resources, in the selection of client groups, and these again are most important

considerations for the success of a program. Process evaluation helps establish and review procedures of instruction, distribution, and organization. Input evaluation and process evaluation help improve things as they are happening. They don't wait until the end to provide feedback. Output evaluation comes at the end of the program or at the end of each important phase of a program. If the other components of the evaluation system--context evaluation, input evaluation, and process evaluation--have been working, there are no unpleasant surprises, at the stage of output evaluation.

Operationalizing Evaluation

Let us now go back to the basic concern of the paper: how to make evaluation operational for a broad impact program of functional literacy or development?

As we have indicated before, it is not possible to think of designing an evaluation program (and then an evaluation system) without analyzing and understanding the design of the development program (or the program system). Again, questions about what data should be collected about whom and on what processes, and procedures and how it should be stored can not even be raised unless the design of an evaluation system has been articulated. Without designing an evaluation system, a data system cannot be designed.[1]

To restate the argument more fully, to develop a functional data system that can be used to provide useful information[2] one must collect and store data

---

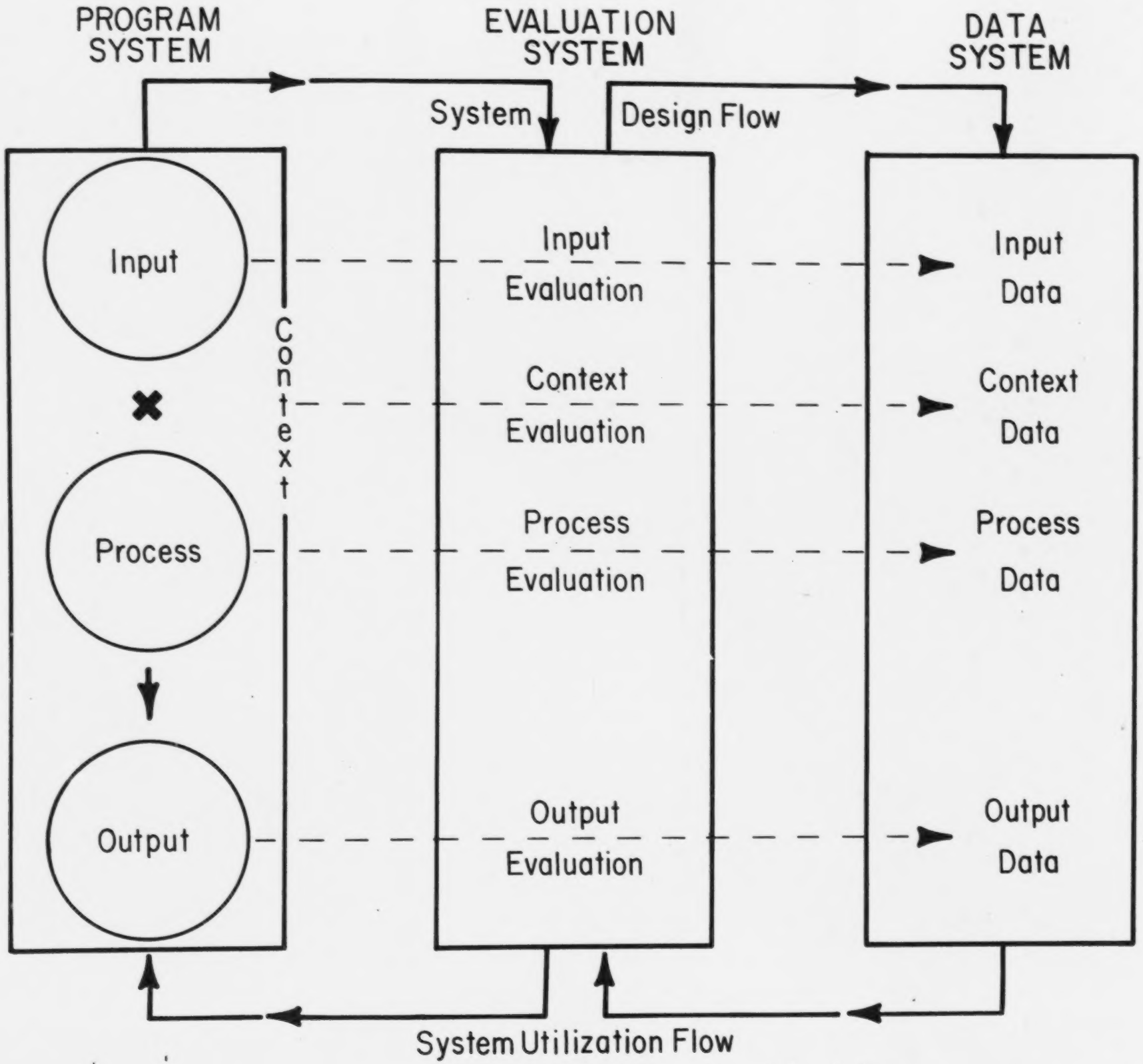
[1]All data systems are not established by evaluators. A data system can be planned independently of an evaluation system when it will not be used for evaluation, but merely as a bookkeeping, and auditing system.

[2]A distinction is sometimes made by specialists in information science between data and information. They suggest that information is what you develop for use from stored data. Data is not information; data is what can be used to produce information for a decision-maker.

with the anticipation that the data collected and stored will have significant usefulness for users. When establishing such a data system, then, we should not be interested in any data or all data on client groups, change agents, instructional materials, settings for teaching, processes, and environment. We should be interested only in appropriate data. What is appropriate data? This question cannot be answered in a vacuum. We must ask: What information will the evaluator need to perform the evaluation tasks and his responsibilities? This shifts our attention to the evaluation system. What this means is that, to build a data system, one must know the objectives, size, and structure of the evaluation system being planned.

Evaluation is rightly defined as the process of obtaining, creating, and providing information to program sponsors, planners, and implementers to enable them to make choices between alternatives. That is, evaluation is the process of servicing decisions. The next question, naturally, is: What decisions? What types of decisions? And decisions about what people, what patterns, what structures, and what operations? These questions, again, cannot be answered in a vacuum. We must have knowledge of the program system to see what actions would be involved, and consequently, what decisions would be involved. We will also have to decide as to what are the important decisions that must be serviced by evaluation based on hard data. Evaluation and information collection cost resources of time and money. Very often, planners may want to have not an ideal evaluation system and an ideal data storage system, but what they can afford. We will discuss this more fully in the section on Evaluation Policy.

To conclude, it should be understood that a data system cannot be planned independently of an evaluation system. In turn, an evaluation system cannot be planned independently of the program system. The figure on the next page presents this idea graphically.


PROGRAM SYSTEM
EVALUATION SYSTEM
DATA SYSTEM
System
Design Flow
Input
Process
Output
Context Interface
Input Evaluation
Context Evaluation
Process Evaluation
Output Evaluation
Input Data
Context Data
Process Data
Output Data
System Utilization Flow

III
THE FIRST STEP IN OPERATIONALIZATION: DEVELOPING A GENERALIZED SYSTEM MODEL FOR THE PROGRAM

As we have indicated earlier, any broad impact program of development and change like functional literacy can be seen as a system, which in turn can be adequately described in terms of four dimensions: inputs, operations, outputs, and context of inputs made, processes used, and outputs obtained.

The first step in an analysis of the Program System for purposes of designing a related evaluation system would consist in developing a generalized system model for the program. That is, a comprehensive description of the Program, in system terms, would need to be developed. Such a description for a functional literacy program would emerge in answering the following questions:

(a) What inputs are being made into the functional literacy project?

(b) What operations and processes are being used within the project to convert inputs into required outputs?

(c) What outputs are intended to be obtained and what might be some unintended outputs? and

(d) What is the socio-cultural context or environment of the program that will interact with and effect inputs, processes, and outputs?

This means that we must list, categorize, and describe variables and treatments in which we are interested as decision makers. Such categorization and listing can indeed be done only if we are clear about the concept on which the program system is based and are clear also about the objectives of the program. If we do not understand the concept or the objectives of the program, we may be listing inappropriate variables.

Let us first try to list the INPUTS in a functional literacy project. We are indeed talking here of a generalized model and hence this listing is also general. Any individual project of functional literacy should find it possible to adapt the list to its specific purposes without much effort.

The inputs, we suggest, may be of three kinds: (A) Clients, (B) Change Agents, and (C) Materials and Facilities.

I. INPUTS

A. Clients

a. Adults
b. Families
c. Community Groups
d. Institutions (Cooperatives, Unions, etc.)
e. Communities and Sub-Cultures[1]

B. Change Agents

a. Literacy Teachers
b. Supervisors
c. Extension Workers
d. Forum Leaders
e. Administrators
f. Technical Specialists and Experts
g. Community Leaders

C. Materials and Facilities

a. Instructional Materials
b. Production Equipment and Supplies
c. Transportation
d. Plants, Spaces, and Housing
e. Duplication and Printing Facilities
f. Broadcasting Facilities
g. Time

Let us go next to a description of another dimension of the system--OPERATIONS AND PROCESSES. At least three categories of operations and pro-

---

[1]Depending upon units of analysis, communities, and even sub-cultures could be considered as inputs. In some cases it might be more useful to consider communities and sub-cultures as Context variables.

cesses may be seen as being conducted within a functional literacy program: (L) Organizational/Structural, (M) Instructional/Formative, and (N) Distributive/Maintenance-Related.

II. OPERATIONS/PROCESSES

L. Organizational/Structural

a. Class Organization
b. Organization for Discussion Forums
c. Organization for Demonstration/Extension Work
d. Organization for Literacy/Development Committees
e. Project Organization
f. Structuring Expert-Counterpart Relationships
g. Inter-Departmental Coordination
h. Institution-Building and Organization Development

M. Instructional/Formative

a. Teaching, Training, and Extension Methodologies for (i) Individuals, and (ii) Groups
b. Pre-Service and In-Service Training Strategies for Staff
c. Methodologies regarding Orientation of International Experts and National Officers from Cooperating Departments at Various Levels

N. Distributive/Maintenance-Related

a. Distribution of Materials and Supplies
b. Transportation and Circulation of Personnel

We may now categorize and list the variables under the OUTPUT dimension of the system. We indicated before that the task of listing inputs, processes, outputs, and contexts for any program system cannot be done independently of the concept in which the program system has been based or without the knowledge of the objectives of the program. The relationship between OUTPUTS on the one hand and OBJECTIVES on the other is direct and should be clearly understood. The outputs of a functional literacy project may be categorized as in the following:

III. OUTPUTS

P. Individuals Who Are:

- Literate
- More Productive
- Healthy
- Active Participants in Socio-Political Institutions
- Efficient Consumers of Information and Education
- Holders of Modern Attitudes Towards Consumption and Investment

OR

Any Alternative Set of Individual Characteristics

Q. Institutions That Are:

- Secular/Modern
- Functional
- Responsive

OR

Any Alternative Set of Institutional Characteristics

R. Communities That:

- Are Cohesive
- Are Healthy and Attractive
- Provide Literate Environment

OR

Any Alternative Set of Community Characteristics

S. Knowledge for the Professional Workers in Literacy:

- On Pre-Conditions for Functional Literacy to Play a Role in Development, and
- On Design, Implementation, and Impact of Programs.

The CONTEXT variables of a system can be similarly analyzed and kept in view to study their impact on the variables of inputs, processes, and outputs.

IV. CONTEXT/ENVIRONMENT

X. Political Context

a. Supportive Factors
b. Neutral Factors
c. Inhibitive Factors

Y. Economic Context

a. Supportive Factors
b. Neutral Factors
c. Inhibitive Factors

Z. Social-Cultural Context

a. Supportive Factors
b. Neutral Factors
c. Inhibitive Factors

With such an analysis of the Program System, we can begin to think in terms of designing an evaluation system.

IV
FROM A PROGRAM SYSTEM TO AN EVALUATION SYSTEM

With a detailed description of the program system in terms of its parameters (inputs, processes, outputs, contexts) and the variables of those parameters (targets, agents, organizational patterns, curricula, social, economic and attitudinal outputs, and environments), we should be ready for the second step--the designing of an appropriate evaluation system.

We have been saying that evaluation is meant to serve sponsors, planners, and implementers of programs in choice of alternatives. In other words, servicing decision-making in all the various aspects of a program is the whole justification for evaluation. The description of the parameters and variables of those parameters will give us an idea of what the decisions will relate to. Our knowledge of theory and our experience with similar work elsewhere will, again, indicate the options for each decision and our information needs to be able to choose between available options. An evaluation system will then need to be designed that will create the information needed to choose between options

available in particular decision settings related to inputs, processes, outputs, and context of a program of action.

It is not useful to think in terms of an ideal evaluation system that might provide all the information that might be needed in all conceivable decision settings. That will be both redundant and expensive. An evaluation system can only be constructed in accordance with an evaluation policy. And, evaluation policies would differ from project to project and from one situation to another depending upon needs and upon resources that might be available.

Evaluation Policy

The policy maker must answer the following questions in regard to evaluation to generate an evaluation policy:

i) Should evaluation be undertaken? (Are the resources available? Is the evaluation effort premature?)

ii) How sizeable should be the evaluation effort if undertaken? (Are the resources available? Which decisions are crucial, which are not?)

iii) Should evaluation be handled internally or externally, or partly internally and partly externally?

The question to be answered first is: Should evaluation be undertaken within a program? There are two parts to this question. One part relates to resources. Another part relates to the history of the program to be evaluated.

Evaluation costs money and requires trained personnel. Funding provided for evaluation is not always additional funding made available to programs. Very often evaluation may be competing with the substantive program for funds. This may be particularly true of programs in developing countries funded from

internal national resources. There the pressure may be for buying books, paying teachers, and hoping for the best, rather than for spending on evaluators and evaluation.

An evaluation policy maker must, therefore, examine his particular situation most carefully. In some cases it may make sense to submit to the program pressures and forget about evaluation even if evaluation might have given the program considerable visibility and prestige. In other cases it may be complete lack of foresight to submit to immediate pressures and not build a fund of tested experiences through systematic evaluation. Where governments may be planning considerable expansions and may most likely invest huge resources in that particular program area in the future, it would make good sense to invest in the evaluation of pioneering, pilot programs.

The second part of the policy question--Should evaluation be undertaken? --relates to the history of a particular program being considered for evaluation. If a program is based on an innovative concept and if program administrators have not yet learned themselves to operationalize the concept into a program, it may simply be premature to evaluate it during the very first cycle of the life of a program. This may be specially true if the new program is competing with an old and conventional program with regard to funding and, consequently, its continuation. What may happen in cases of premature evaluation, is this. The innovative, but new, program, may show no results, or may show comparatively smaller results, not because it was a theoretically insupportable and practically unfeasible program, but because the program had not even been properly engineered for a field test and had not, therefore, delivered its full and latent potential in the form of impact on individuals and communities. On the other hand, an old and conventional program may

show some results because it had been operationalized and could move a little bit, but a little bit that could be demonstrated. The implications of premature impact/output evaluation should be obvious from the preceding discussion. It does not mean, however, that program planners should not collect both impressions and data for developing operational insights during the first cycle of a pilot program to apply it in a second cycle when more systematic and comprehensive evaluation systems may have to be designed.

A second policy question relates to the size or comprehensiveness of the evaluation system to be designed. Again, part of this question is a question of resources which we have discussed above. The second part of this question relates to the nature of decisions. How crucial and far-reaching are the decisions? What will be the costs or risks involved if a set of decisions are made on the basis of experience and impressions (and not on the basis of information made available by the evaluation system) and those decisions turn out to be wrong? Can we divide decisions in clusters and arrange them on a continuum of their risk-values?

Should evaluation be handled internally or externally or partly internally and partly externally? This is the third policy question. Different answers are available, but we suggest here the one that we think is the most reasonable. We suggest that the context evaluation, input evaluation, and process evaluation should generally be handled internally. Some of the problems of input, process, and context evaluations may, however, be too tough to handle in terms of design and analysis and may be handled in cooperation with outside experts. Also, all internally generated data and documentation should be available to external evaluators when these external agents undertake any evaluation. Impact evaluation might best be handled externally. This should not, however, mean that impact evaluation should be out of bounds for internal evaluators and program teams.

As part of evaluation policy, it would also be necessary for the policy-maker to both understand and in turn to clarify for others that evaluation is not done only by evaluators. Evaluation is a cooperative activity jointly and cooperatively handled by evaluators and program implementers--trainers, curriculum specialists, audio-visual experts, extension workers. In this relationship, each does what he can do best. The program specialist defines problems, defines alternatives, details variables, and states the questions he would like to get answers to. The evaluator works with the specialist in the process of definition of problems and alternatives, selection of appropriate variables and in statement of questions. In this stage of cooperation, the initiative must be with the program specialist. The evaluator then helps in the problems of design, instrumentation, data collection, and analysis. At this time, the evaluator takes the initiative.

Evaluation System

Translating an evaluation policy into an evaluation system is the next step. At the system level there are a different set of questions that will need to be answered. These questions are:

1) What would be the intended outputs of the program that will be selected for inclusion in an evaluation system? And, what might be some unintended outputs?

2) What will be the social/physical units which will be studied as carriers of these impacts?

3) What would be the indicators used to demonstrate that an intended impact is indicated in any social unit--individual, group, institution, or subculture? These, because of the types of changes we are looking for--new atti-

tudes, new orientations, new social climates, new organizational patterns--are not available for direct observation.

4) At what intervals of time would variations in indicators be recorded, both on long-term and short-term bases?

5) What inputs/interventions, what processes, and what settings/contexts will be given special attention within the context of the objectives of the total program?

6) What evaluation designs will be used to generate needed information, especially about processes?

7) What will be the division of labor between evaluators and program specialists and what will be the patterns for taking initiatives and implementations?

8) What will be the points of data collection, data collation, data storage, and data analysis? (Answer to this question will determine what type of data system would ultimately emerge.)

Of course, it would be necessary, for establishing such an evaluation system, that the problems of definition of terms be taken care of for the purposes of establishing inter-area or inter-project communication and comparability.

Let us ask the questions listed above in regard to the functional literacy program and suggest some <u>tentative</u> answers. The intended outputs of a typical functional literacy program have already been listed above. A particular project or program, as we have already indicated, may not be in a position to measure <u>all</u> the outputs possible. It may limit its concerns only to two or three or four important outputs, for example, literacy, productivity, changes in group perceptions, establishment of modern economic institutions, etc.

Impact is on some units--physical (agricultural tools, clothing, living spaces, fields) or social (individuals, families, communities). A functional literacy team must decide, for the purposes of establishing an evaluation system, as to what units will be studied as targets and absorbers of impacts. A hypothetical project may consider as targets, cotton growing, male farmers between 20 and 45 and no other adult group in the communities. Other literacy workers may also include in their study, wives of the adults considered as the target group. Homes may not be studied, but fields may be. Churches may not be studied, but Credit Cooperatives may be. A selection of units to be studied as carriers of the program's impact, is thus necessary.

While physical impact is often available for direct observation, social-psychological impact (on individuals, groups, and institutions) has often to be indirectly inferred. Something observable must indicate a change that cannot be directly observed. We cannot see fever, but a thermometer indicates it. We cannot see intelligence, but a high or low score on an intelligence test might indicate the quality of intelligence of a person. It is part of building the evaluation system to decide as to what would be the indicators of the impact that is sought to be demonstrated. Further, one condition (of impact on a social unit) need not be studied by one indicator. More than one indicator may have to be used. Again, what observable indicator would demonstrate what unobservable inner condition, is not always clear to social scientists. Different social scientists may suggest different indicators of an inner condition. More wisely, a complementary set of indicators may be suggested. These decisions are part of building an evaluation system. (It should be noted that these are also the steps in building a sensible program.)

The Unesco functional literacy program has taken some of the following indicators[1] of impact in view:

---

[1]For details, see Unesco documentation on indicators, EDA/1377/0471, revised report, 1971.

1) Acquisition and use of the skills of reading, writing, arithmetic (Test Scores, Observation)

2) Occupational/technical knowledge (Scores)

3) Socio-economic knowledge (Scores, Observation)

4) Knowledge of health, hygiene, and nutrition (Scores, Records)

5) Attitudinal change with regard to modernization (Scores, Records)

6) Adoption of recommended practices (Observation)

7) Scholarization of children (Records)

8) Productivity and income (Scores, Records)

9) Change in consumption patterns (Observation)

At what intervals would indicators be measured or observed? First, there is the question of base-line surveys. Such surveys are important to conduct if any statements regarding impact and change have to be made in a before and after format. Second, this would involve questions of phasing the program and of determining intermediate objectives for each phase. Third, there is the question of long-term impacts after two, five, ten, or twenty years. How far do we go?

Discussions in the immediately preceding paragraphs have related mainly to questions of impact, that is, only to output evaluation. An important concern in building an evaluation system is, however, related to inputs, processes, interventions, and settings of such interventions. The team working on building an evaluation system must select crucial processes, interventions, and their settings, and variables in inputs that must become part of the input, process, and context evaluations. These input variables and settings must then be suitably differentiated. Also, the varied interventions must be standardized to make conclusions about what intervention resulted in what impact, in what setting, and through what process.

We have also pointed out above that social psychological phenomenon cannot always be studied by direct observation. Observable indicators have to be used to demonstrate the possible existence of an inner condition of a social unit. To this we must add that evaluating involves, most often, comparisons in status of impact-carriers before and after interventions; comparisons between groups that had or did not have some intervention or sets of interventions; or comparisons between groups that received similar interventions in differential settings undergoing different processes. This means that to generate all this information evaluation, <u>designs</u> will have to be used. Part of the design problem is technical, but part of it is a problem of planning the evaluation system itself and must be considered early at the stages of designing the evaluation system.

Evaluation would fail by default if distribution of labor between external and internal evaluators, on the one hand, and between internal evaluators and program staff, on the other, is not clarified and patterns for responsible initiatives and implementation are not established. It is also part of the problem of building an evaluation system to make decisions about who will collect what data, where this data will be collated, and stored, and who will analyze it for what purpose.

We can now move to our next concern--Building a Data System to serve the evaluation needs.

V
DESIGNING A DATA SYSTEM

Once the outlines of an Evaluation System have emerged, the next step is to create a Data System. The designing of the evaluation system and the data

system are seen as two separate steps only for analytical purposes. The two design tasks will infact be undertaken at the same time and the development of the data storage system will indeed contribute to the design of the evaluation system itself.

There is some information and data that a program will leave in its wake as it is implemented, for example, information on annual budgets, number of teachers employed, number of adults in the program, attendance figures, etc. The problem here is to record all this data in a suitably designed register or in a table or on cards. All the data that a program will generate will not be numerical. Some of it will be in the form of documentation, such as, minutes of meetings, files, teacher and supervisor journals, log books, specimens of primers, sets of instructional materials, class schedules, supervisors' itineraries and reports, pictures, broadcast scripts, and such. This type of data also must be carefully recorded and filed before it is lost. This is certainly an important component of any Data System.

The data that is generated just by the fact of the implementation of a program is not likely to be sufficient data for purposes of evaluation. Some additional data will have to be especially generated. For generating this special and needed data, we will use some Evaluation Design. That is, we will use the methodology of designing experiments or evaluations. We will also make tests and instruments, use comparable groups, and samples, and so on. How to design evaluation studies? How to draw appropriate samples? How to make tests, questionnaires, and interview and observation schedules? How to treat data statistically? How to interpret data to generate information? All these are technical questions that must be dealt with separately in a subsequent paper. In this paper we must confine ourselves to the problems of operationalization of evaluation systems and to the design of related components and systems.

A Data System for Functional Literacy must have at least two components and each component must have two types of data as shown below:

	Data Generated in Implementation	Data Specially Created
Numerical Data		
Data included in Documentation		

The criteria in building a Data System are the following:

1) All needed information may be collected with the greatest economy of effort. Useless information may not be collected. Duplication in information collection may be avoided unless a replication of an experiment is underway or a reliability check is being applied to data.

2) While considerable amounts of information will have to be especially generated at short notice by evaluators working on programs, a good Data System should be able to <u>anticipate</u> most of the information needs of the program.

3) The data should be stored for easy retrieval and use. Data should be <u>disaggregated</u> enough so that it can be aggregated by the evaluator himself in various ways to suit his particular purposes. If scores on attitudes towards own education and education of children are available separately, in disaggregated form, an evaluator may be able to treat the two scores separately for cre-

ating suitable information. If these two scores are already aggregated, the evaluator may find it impossible to disaggregate them into two separate scores.

4) The data system should be designed for flexibility. It should be possible to extend it with comparative ease or to adapt it to new data needs as they emerge.

The evaluation system when designed will tell us what different variables of program system will be studied. The evaluation system will also lay down what will constitute evidence of a process taking place, and what will be the indicators of impact on individuals, groups, and communities. As we have indicated in the previous section, developing agreements on indicators and on the nature of admissible evidence are an important part of developing an evaluation system.

A dimension that must be clarified before constructing a Data System is the temporal dimension. At what points in time will measurements be made to generate data? In cases where program activities fall into natural cycles and phases, the time question is answered for us. In other cases, however, decisions on time schedules will have to be specifically made.

Let us relate the discussion so far to the Unesco Experimental Functional Literacy Program. It can be asserted that most or a large part of data needed for context evaluation, and process evaluation may have to be especially generated through evaluation design. Only some of it may be generated as part of program implementation. On the other hand, most or a large part of data needed for input and output evaluation may be generated through the implementation of the program itself. Though some input and output data may have to be specially generated through evaluation design. We argue in the following that building a data system is mostly (though not completely) a matter of generating and storing for

retrieval, data about four configurations[1]--Individuals, Groups, Institutions, and Sub-Cultures. We will indeed need data, additionally, on artifacts and tools used in the program. However, we are suggesting that whether we are making context evaluation or process evaluation, input evaluation, or output evaluation, we will be concerned with data on four social configurations, to repeat: Individuals, Groups, Institutions, and Sub-Cultures. Inputs in a functional literacy program are individuals, groups, and institutions; in some cases whole sub-cultures. Input evaluation thus must collect data on these configurations (and, of course, additionally on artifacts, tools, and material inputs). Outputs in a functional literacy program are, again, these configurations which have now been changed because of the program, in a desirable direction. In output evaluation also, therefore, while studying impact we are basically concerned with these configurations (and about some physical/material entities like seeds, agricultural tools, homes, and fields). Again, processes can only be studied through what these processes process; that is, by observing what they do to those to whom these processes are applied. The processes of instruction, formulation, extension, distribution, management, and organization, can thus be studied only in terms of what they did to the social configurations involved in a total broad-impact program. (Some process, of course, may be applied to physical entities, fields, and homes). Again, in context evaluation, we are interested in the social, economic, and political presses on the environment and these are determined by sub-cultures within which a broad impact program is being conducted and by the

---

[1]These are the four basic social configurations, and all change events can be analyzed in terms of configurational relationships between one or more of these configurations. In a broad impact program like functional literacy, all the four configurations will most often be involved. See H.S. Bhola, "Configurational Theory of Innovation Diffusion," Indian Educational Review, Vol. 2, No. 1 (January, 1967). Again, H.S. Bhola, "The Methods and the Materials of Functional Literacy," Literacy Discussion, Vol. 1, No. 4 (Autumn, 1970) includes a brief discussion of the configurational model of change.

institutions located in those sub-cultures. Here, again, though physical and material components of the environment will have to be taken into consideration, we are dealing with four basic social configurations.

To summarize, whatever the evaluation type (context, process, input, or output) it will need data on four social configurations--Individuals, Groups, Institutions and Cultures or Sub-Cultures.. Additionally, it will need information on tools, and techniques, and artifacts that these configurations may be working with in their development, but data on social configurations is basic.

The following social configurations might then be subject of study for data collection, depending upon the objectives of various programs:

1. Individuals
   - Farmers
   - Teachers
   - Local Leaders
   - Supervisors
   - Extension Workers
   - Administrators
2. Groups
   - Literacy Classes
   - Radio/Discussion Forums
   - Demonstration Groups
   - Traditional Groups
3. Institutions
   - Schools
   - Cooperatives
   - Local Courts
   - Party Cells
   - Families
   - Factories
   - Offices/Departments/Ministries
4. Sub-Cultures
   - Communities
   - Factions
   - Castes

To sum up, a data system will include both numerical information and documentation. Some of this numerical data and documentation will become available by the sheer fact of implementing a program if any records are kept at all. Some data and documentation will have to be especially generated by using eval-

uation design and methodology. Again, most data will relate to the four social configurations--individuals, groups, institutions, and sub-cultures involved; and some, of course, to the tools, techniques, and artifacts that those configurations worked with to create development.

The next important step in data system design will be to design data storage formats such as cards, tabulations, forms for recording data on selected configurations--individual farmers, groups, families, cooperatives, communities--which can be used to store data on various attributes of these configurations at various intervals of time; and to establish filing systems and a morgue for documentation.

It is not possible within the scope of this paper to deal with the questions of what these cards, tabulations, and forms may look like and what might be a system for filing and storing documentation. We might deal with these questions separately in a subsequent paper.

VI
EVALUATION--NOT BY EVALUATORS ALONE

Evaluation is not something which is done by evaluators alone and by no one else. We have emphasized repeatedly that evaluation is a group task. It is a group task all the way through. Not only should evaluation be planned by a whole group working on a project, it should also be conducted by the whole group, each member contributing what he is best prepared to contribute to the evaluation effort.

If evaluation is, indeed, as we have defined it, a process of servicing decisions, that is, of serving decision-makers, how could it be handled without the decision makers, by the evaluator working in isolation? Unless the program team

sits together to analyze the program, details the program parameters, lists the variables involved, anticipates likely decision settings, and articulates the information needs for making those decisions, how is the evaluator going to do what he is supposed to do? At this early stage of the process of program analysis, the evaluator listens, asks questions, but does not answer them for the group. He does not tell what is worth evaluating and what is not. That initiative must lie with program specialists. If program specialists do not take the initiative at this stage of work, they probably do not understand the program or its structure and will be unable, therefore, to operationalize it, much less be able to assist in the development of an evaluation system.

The evaluation policy for a program is, again, something that has to be designed not by the evaluator, but by the program group working together. Of course, the evaluator will participate in the process of policy making and will clarify the various possibilities in the context of available resources and established priorities.

At the stage of design of an evaluation system and the design of a data system, the evaluator may take all the possible initiatives, but even here he should involve the program specialists. A lot of data, as we have indicated, will be generated through sheer implementation of the program, and this is data which will be used both by evaluators and program specialists. Storage formats should, therefore, be designed that can serve both purposes; and they can be designed to serve both purposes if both the evaluator and program specialists contribute to the design of the evaluation system and the data system.

In evaluation methodology, in the preparation of questionnaires, observation inventories, and interview schedules, the evaluator will play a prior part because he is technically trained to perform those tasks. However, in data collection,

again, he should not be the only man. The whole team should participate because data collection and feedback is everybody's business, it is part of all sensible program activity. Field workers, supervisors, and area administrators can collect data only with a little more effort on their part; and a parallel bureaucracy should not be created by evaluation merely for data collection. Also, information created by evaluation must be used by program implementers and they will have more confidence in the information they themselves helped create. Lastly, they will grow as professional workers through participation in evaluation.

Operationalizing evaluation is a tough job. Since it must begin with questioning the program, must ask hard questions about program objectives, its theoretical and value assumptions--which are not always clear--and must insist on consistency in how inputs are supposed to be related to processes to produce required outputs within a particular context, it is likely to generate considerable trauma among program specialists. Operationalization of evaluation thus can not promise always to be fun. For program specialists who did not raise the right question or left too many of those unanswered--for lack of ability and sense of security--building an evaluation system can be a terrible experience.

VII

In our previous discussion on the operationalization of evaluation, we have indicated how a data system must reflect the needs of an evaluation system which in turn must respond to the decision-making needs of a program system.

Systems are, of course, composed of sub-systems and are themselves parts of meta systems. To take the example of the Unesco/UNDP Experimental World Functional Literacy Program each country project could be considered to be one pro-

gram system. However, a single country project might be working with different client groups, and using different processes of instruction and formulation to produce particular types of outputs in different socio-economic contexts. That is, country projects as systems may be composed of sub-systems. Again, the total Unesco program comprising twelve or more country projects could be considered a meta system with individual projects in different cultural contexts as its sub-systems.

A meta evaluation system that would reflect the decision making needs of the total program system, such as Unesco Experimental Functional Literacy Program, to be operationalized would require our going through the same steps we have discussed in detail in the preceding. It cannot be left to emerge of itself from its various parts; but indeed is a complex exercise in program definition, clarification of decision making needs, and thereby of the needs for specific data at some central point.

For a meta evaluation system, again, we will have to begin with a generalized model of the total program system. Since such a meta evaluation system would deal with many country projects, the program model will have to be generalized at an even higher level of abstraction. What kinds of decisions will be made by the decision makers at the meta system level? That would be the next question to answer. Once it is clear what kinds of decisions will be made at this level, the next steps regarding evaluation policy, evaluation system design, and data system design, can be taken. These problems have already been discussed in great detail in the preceding, but some special problems faced in creating such a meta evaluation system can be indicated here. These arise from the need to develop a strong central "nervous system" to communicate needs, receive data and monitor breakdowns; the requirement to develop uniform vocabulary

and to take care of problems of definition; and the need, again, for developing instrumentation and procedures for interpretation so that generalizations could be made across projects, and across cultures.

All these tasks are by no means easy, but certainly can be accomplished. Only those questions are impossible to answer that are never raised!

VIII
SUMMARY

In this paper we have clarified the concept of evaluation, contrasting it with research and supervision, and have distinguished four types of evaluation: Context evaluation, input evaluation, process evaluation, and output evaluation. We have applied these concepts to the Unesco program of functional literacy to bring out some of the special problems of building evaluation systems for broad-impact programs.

As the Chinese box, in the box, in the box cannot be opened without first opening the box, outside the box, outside the box; similarly a data system cannot be built nor data collected and stored without first building an evaluation system; and an evaluation system cannot be conceived without first analyzing and understanding the program system and its decision needs. The need for an evaluation policy has also been brought out indicating that there is no such thing as a complete evaluation system, but that each project must build its own evaluation system in terms of its needs, priorities, and resources.

We have emphasized repeatedly the fact that evaluation is a task for the whole group with the evaluator playing a special technical role for which he is best prepared.

We have in this paper kept our interests confined to the problems of making evaluation <u>operational</u> in the context of a broad impact program. The new method-

ology of evaluation and some of the tools and techniques that are now available to the evaluator have not been discussed. Nor have we suggested specimens of storage devices and formats which could be used to store numerical data and documentation. These must be discussed separately at some later time.

ERIC Clearinghouse
JUL 10 1972
on Adult Education